

On-Line Estimation of LO Noise and Optimisation of Servo Parameters in Atomic Clocks

Ian D. Leroux,* Nils Scharnhorst, Stephan Hannig, Johannes Kramer, Lennart Pelzer, Mariia Stepanova, and Piet O. Schmidt

QUEST Institute for Experimental Quantum Metrology, Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany

(Dated: 25th January 2017)

For atomic frequency standards in which fluctuations of the local oscillator frequency are the dominant noise source, we examine the role of the servo algorithm that predicts and corrects these frequency fluctuations. We derive the optimal linear prediction algorithm, as well as practical procedures for optimising the parameters of a conventional integrating servo and characterising the flicker and random-walk noise of the local oscillator without interrupting the normal operation of the standard. In particular, we show that near-optimal servo parameters can be calculated without independent knowledge of the local-oscillator noise, based only on monitoring of the error signal. Using simple analytical models and numerical simulations, we establish optimum probe times as a function of clock atom number and of the dominant noise type in the local oscillator. This leads to an argument for the superiority of methods that increase the measurement signal-to-noise ratio (increasing atom number, spin squeezing) over those that increase the effective transition frequency (maximally-correlated states) when trying to improve the stability of local-oscillator-limited frequency standards.

The instability of frequency standards limits the total uncertainty achievable in a measurement of finite duration [1, 2]. This limit can be practically relevant even when performing measurements of static frequency ratios, since many-month-long measurement campaigns place stringent demands on the reliability of all components in an experiment. Instability becomes a fundamental concern when attempting to measure time-varying frequency ratios. For instance, in the emerging field of chronometric leveling [3–5], direct observation of tidal fluctuations expected in the gravitational red shift [6] requires frequency ratio measurements with a fractional uncertainty at the level of 10^{-18} to be completed in a matter of hours. Physics beyond the Standard Model might be detectable in clock frequency ratio measurements as postulated transient shifts associated with dark-matter domain walls [7] or ultralight scalar dark-matter candidates [8, 9]. Searches for such signals would require even faster measurements, completed in seconds. At such time scales, the statistical uncertainty due to instability plays a far greater role than long-term systematic uncertainty in current standards.

Of the noise processes contributing to the instability of atomic frequency standards, the most fundamental one is quantum projection noise [10], which arises from the discreteness in the measurement results obtainable from a finite number of atoms. For an ensemble of N uncorrelated two-level atoms, this noise imposes a minimum statistical uncertainty

$$\Delta\phi_{\text{QPN}} = \frac{1}{\sqrt{N}} \quad (1)$$

on any measurement of the phase accumulated in an atomic superposition state. For a standard operating at a frequency ω and in the ideal case of dead-time-free Ramsey interrogation without technical noise, this leads to a long-

term fractional instability [11]

$$\sigma_C(\tau) = \frac{1}{\omega T \sqrt{N}} \sqrt{\frac{T_c}{\tau}} \quad (2)$$

where T is the duration of a single Ramsey interrogation and T_c is the length of the frequency standard's operating cycle, such that τ/T_c measurements can be performed in an averaging time τ . This quantum projection noise limit (QPN)¹ for clocks using uncorrelated atoms depends on the experimenter's choice of probe time T , becoming arbitrarily small for sufficiently long probe times. Thus, Eqn. 2 sets no limit on achievable clock instability at long averaging times unless some additional scale in the problem restricts the choice of T .

One such restriction is set by excited-state decay in the atoms, which sets a fundamental limit to interrogation times. The performance of optical frequency standards operating at this limit has been analysed in Refs. [12–15]. However, for many of the optical frequency standards now being investigated, frequency fluctuations of the local oscillator restrict T to less than a second even when the atoms' excited-state lifetime is measured in minutes (^{87}Sr) or even years ($^{171}\text{Yb}^+$) [1]. Because the local oscillator's noise is common to all the atoms in the standard, and because it typically exhibits significant power-law temporal correlations, its effects are qualitatively different from those of excited-state decay. In fact, it might at first glance seem odd that local-oscillator noise limits clock stability at all: the local oscillator frequency is in some sense the measurand in an atomic frequency standard and its fluctuations are constantly monitored and corrected. Local-oscillator noise affects the stability of the standard only to the extent that it cannot be corrected by feedback from the atoms. This can happen, for instance, if the cyclic atomic interrogation protocol allows undetected aliased frequency components of

* Ian.Leroux@nrc-cnrc.gc.ca

¹ sometimes referred to as the Standard Quantum Limit (SQL)

the local-oscillator noise spectrum to contaminate the output signal of the standard, a phenomenon known as the Dick effect [16]. Even in the absence of the Dick effect, however, the quantised measurement signal from the atoms has a fundamentally limited dynamic range: one cannot extract more than $\log_2(N+1)$ bits of frequency information from a single measurement of N atoms [17]. The useful domain of the measurement, i.e. the frequency band in which it can be unambiguously interpreted, must be broad enough to cover the frequencies which the local oscillator is likely to emit in the interrogation. Frequency excursions beyond the domain for which the reference provides useful information lead to less informative measurement results, and hence to degraded instability. In the worst case the servo, working from ambiguous or uninformative measurement results, may be unable to keep the output frequency locked to the atomic reference. The output frequency then either hops between different zero-crossings of a frequency-periodic Ramsey error signal or drifts aimlessly far from the resonance of a Rabi error signal. This case is catastrophic and the operating parameters must be chosen to make it vanishingly unlikely. Thus, even in the absence of the Dick effect (e.g. with dead-time-free Ramsey interrogation [18]), the achievable measurement resolution ultimately depends on the scale of local-oscillator frequency fluctuations seen by the atoms, and hence on the performance of the clock's feedback loop which corrects these fluctuations.

In this work, we study the limits to the stability of frequency standards dominated by local-oscillator noise with realistic temporal correlations. We focus on clocks using a single ensemble of atoms periodically interrogated using the same protocol for every interrogation cycle, whose instability we quantify using the Allan variance at long times. Our work is thus less general, but more directly relevant to current experiments, than analyses of multi-ensemble clocks or of interrogation protocols which are modified on-the-fly [19–22], and our approach is a more concrete complement to the derivation of universal performance bounds in mathematically idealised settings [22, 23]. Using simple analytical arguments and numerical simulations of clocks with different local-oscillator noise spectra, we study the performance of the servo controller which predicts and corrects local-oscillator noise and then analyse its implications. After establishing notation and conventions in Sec. I, we begin by deriving the optimal linear prediction algorithm and evaluating its performance in Sec. II. In Sec. III we show that a feedback controller with near-optimal performance can be designed without prior knowledge of the noise spectrum, by monitoring the error signals in normal clock operation. We also show that the same techniques provide useful diagnostic information on the local oscillator's noise, allowing on-line monitoring of its performance. We turn to the effects of the noise in Sec. IV, in which we derive a modification to the QPN that takes into account the performance of the servo controller. This modified QPN formula predicts an overall limit to achievable clock instability, which is attained for an optimal choice of atomic interrogation time that we discuss in Sec. V. Sec-

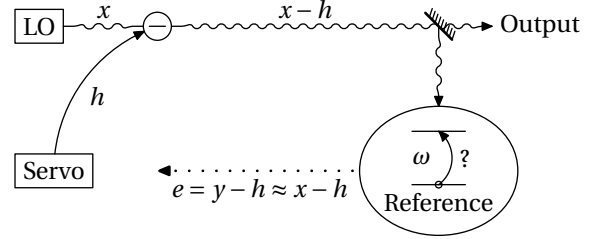


Figure 1. General model of a periodically-stabilised atomic clock. A local oscillator (LO) emits a signal with a fractional deviation x from the nominal clock frequency. The servo controller attempts to predict x and corrects the frequency by its prediction h . The corrected signal (with fractional frequency deviation $x-h$) is then used to interrogate an atomic reference, which produces an estimate e of the prediction error or, equivalently, an estimate y of the LO's unknown frequency deviation x . These estimates can be used by the servo in future predictions.

tion VI considers the merits of using entangled atomic states to modify the phase resolution of Eqn. 1, giving a simple dimensional argument for the disappointing performance of maximally-correlated states in atomic clocks and arguing for the superiority of states that enhance the dynamic range of atomic measurements [17, 24–26]. This result is complementary to that of Ref. [12], which considered independent dephasing of the atoms rather than the collective dephasing associated with the LO, and takes into account temporal correlations in the LO noise rather than assuming a white spectrum as in Ref. [20, 27]. Section VII considers the instability of the clock at short times, which may be limited by finite feedback gain rather than measurement noise, showing that the second integrator recommended in Ref. [15] to correct for linear drift of the LO is also necessary to saturate the QPN limit in the presence of random-walk noise. We conclude with some remarks on proposed frequency standards whose design does not follow the conventional pattern we consider in this work.

I. SETUP AND NOTATION

Figure 1 sketches the structure of the frequency standards we consider and summarises the notation we will use for the various signals we must consider. The goal of the standard is to produce a continuous classical oscillatory signal whose frequency corresponds to a reference transition frequency ω in a particular atomic species. As the classical signal is generated by a macroscopic local oscillator (LO) subject to environmental perturbations, its frequency ω_L will differ from the target frequency ω by a fluctuating fractional discrepancy $x = (\omega_L/\omega) - 1$. The scale of these fluctuations is summarised in the Allan deviation $\sigma_L(\tau)$ of the LO. In order to suppress these fluctuations, a servo controller generates a prediction h of the LO frequency error, which is used to frequency-shift the LO output signal back to atomic resonance. The resulting signal, with net fractional frequency error $x-h$, is provided to the users of the standard and has

an Allan deviation $\sigma_C(\tau)$. This corrected signal is supplied to a reference, where it interacts with N atoms according to some fixed interrogation protocol, such as Ramsey or Rabi interrogation. Measurement of the atoms' state at the end of the interrogation protocol conveys some information on the residual frequency error $x - h$, which we express as an error estimate e . The error estimate might, for instance, correspond to the imbalance of atomic state populations at the end of a Ramsey sequence divided by the accumulated phase ωT . For consistency, we express the error estimate in the same units as x and h , so that $y = h + e$ is an estimate of the LO's uncorrected frequency error x , one that uses only the most recent atomic data and takes no account of previous measurements. Note that y and e differ from, and fluctuate more than, x and $x - h$ respectively, because they are affected by the noise of the atomic reference.

We consider periodically-stabilised frequency standards with an operating cycle of period T_c , where the reference provides a series of error estimates $\{e_i\}$. At any given point in time, we label them as follows: e_1 is the most recent available error estimate, e_2 the preceding error estimate, and so forth. e_0 is then the error estimate which will be produced at the end of the current operating cycle. We label the other signals similarly: h_0 is the servo's prediction of the (average) LO frequency error x_0 during the current operating cycle, while h_j and x_j correspond to the j th most recent completed cycle. Causality requires that the servo compute the prediction h_0 without knowledge of e_0 , using only $\{e_1, e_2, \dots\}$ or, equivalently, $\{y_1, y_2, \dots\}$.

The Allan deviation $\sigma_C(\tau)$ is that of the physical signal produced by the frequency standard as it is operating. With the exception of Sec. VII, most of the analysis presented in this work also applies to "paper clocks", i.e. virtual signals generated by post-processing measurement data. Although the post-processing need not respect causality and can use later measurements to correct estimates of the frequency at earlier times, the quality of the measurements themselves still depends on the ability of the (causality-respecting) servo to keep the corrected LO frequency near the atomic resonance frequency ω while the clock is running, and constraints on this ability affect the performance of the reference no matter how the resulting data is subsequently used.

Where it is necessary to assume a definite interrogation protocol in the atomic reference, we will focus on dead-time-free Ramsey interrogation, where the measured signal depends on the average of the corrected signal frequency during some interrogation time T . While we assume in our examples that T , which sets the frequency resolution of the interrogation, is equal to T_c , which sets the repetition rate of the interrogation cycle, the two times are conceptually distinct and we will use separate symbols for them throughout. References whose operating cycle includes dead time ($T < T_c$) or which use a different interrogation protocol (such as Rabi or hyper-Ramsey [28, 29]) will suffer from the Dick effect, which can be modelled as additional measurement noise in the atomic reference.

In numerical examples we will consider LOs with simple

power-law noise, such that $\sigma_L^2(\tau) \propto \tau^\mu$, with $\mu = -1, 0, 1$ for white frequency noise, flicker frequency noise and random walk of frequency noise, respectively. As argued in the introduction, the LO noise gives the problem a characteristic time scale which ultimately limits the useful resolution of measurements on the atoms. We define this time Z , without assuming a particular form of LO noise spectrum, by the implicit equation

$$\sigma_L(Z_c)\omega Z = 1 \text{ rad}, \quad (3)$$

where Z_c is the cycle time of the clock when operated with a probe time Z . In other words, Z is the choice of probe time for which the LO Allan deviation at one clock cycle is as large as the quantum projection noise of a single atom (Eqn. 2 with $N = 1$). This definition lets us combine the LO noise and the choice of probe time into a single dimensionless parameter T/Z which can be compared between clocks of different types using LOs with different performance. Note that Z will be on the order of a few seconds for a typical current optical frequency standard with a fractional LO instability around 10^{-16} .

II. SERVO CONTROLLER DESIGN

We now focus our attention on the servo. Given a history \dots, h_3, h_2, h_1 of its own past predictions and of the corresponding error signals \dots, e_3, e_2, e_1 obtained from the atomic reference, it must make a prediction h_0 of the LO frequency in the next operating cycle. The prediction should take into account the temporal correlations of the LO noise, which dictate the timescale over which past measurement results remain relevant to predicting future LO behaviour.

We begin by considering the simple integrator, the basic building block of the servo algorithm used in most contemporary optical frequency standards [15]. In our notation, the simple integrator makes the prediction

$$h_0 = h_1 + g e_1 \quad (4)$$

where g is a dimensionless gain specifying the fraction of the frequency error measured in the last cycle to apply as a correction to the last prediction. The prediction can also be expressed in terms of past estimates of the LO's fractional frequency deviation $\{y_k\}$ as follows:

$$h_0 = \sum_{k=1}^{\infty} g(1-g)^{k-1} y_k. \quad (5)$$

While Eqn. 4 is easier to implement, Eqn. 5 is easier to reason about because the statistical properties of the estimated LO frequency y are mostly determined by the LO noise and by the measurement noise of the reference, depending only weakly on the design of the servo controller itself. To a good first approximation, then, we can take the fluctuations and correlations of the $\{y_k\}$ as given, and try to choose g so as to minimise the error of the prediction h_0 .

It is instructive to study the broader class of linear predictors, whose predictions are weighted averages of past LO frequency estimates of the form

$$h_0 = \sum_k w_k y_k, \quad (6)$$

where the weights w_k are required to satisfy the normalisation condition

$$\sum_k w_k = 1. \quad (7)$$

The simple integrator of Eqn. 5 is a special case of a linear predictor, with $w_k = g(1 - g)^{k-1}$. The optimisation of such linear predictors has been studied extensively since the pioneering work of Wiener [30] and Kolmogorov [31] (see e.g. Ref. [32]). Here we derive the minimum-mean-squared-error predictor in a form similar to that used for ordinary kriging in geostatistics (see e.g. [33]). We begin by computing the mean squared difference between the prediction h_0 and the next frequency estimate y_0 :

$$\langle (h_0 - y_0)^2 \rangle = \left\langle \sum_j \sum_k w_j w_k (y_j - y_0)(y_k - y_0) \right\rangle \quad (8)$$

$$= \mathbf{w}^T \mathbf{C} \mathbf{w}, \quad (9)$$

where we have collected the weights $\{w_k\}$ into a vector \mathbf{w} and introduced the two-sample covariance matrix for the estimated LO frequency, whose entries are defined as

$$C_{jk} = \langle (y_j - y_0)(y_k - y_0) \rangle. \quad (10)$$

Note that $\langle (h_0 - y_0)^2 \rangle$ is not the same as the mean squared prediction error $\langle (h_0 - x_0)^2 \rangle$, since it also includes the noise of the atomic reference which estimates that error. Provided, however, that the atomic reference is unbiased, the same choice of weights will minimise either measure of noise, so we proceed to minimise Eqn. 9 and find that the optimum weights satisfy

$$\mathbf{C} \mathbf{w} = \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix} \quad (11)$$

with λ a Lagrange multiplier that must be chosen to satisfy the normalisation constraint of Eqn. 7. Thus the optimal weights can be found by solving Eqn. 11 for \mathbf{w}/λ and normalising the result, provided that one knows the covariance matrix \mathbf{C} . If the noise properties of the components in the frequency standard are known, then \mathbf{C} can be computed simply as the sum of matrices for each independent noise process. Explicit expressions for the \mathbf{C} matrix associated with a known noise spectrum are provided in Appendix A. As discussed in Sec. III, \mathbf{C} can also be estimated, and the servo controller optimised, *without* prior knowledge of the system noise properties, using only data generated during normal clock operation.

Although linear predictors with arbitrary coefficients are not difficult to implement following Eqn. 6, one can also use the preceding formalism to optimise the gain of conventional integrators. Appendix B derives an explicit, albeit cumbersome, formula for the optimal integrator gain given known noise model parameters. Alternatively, one can use Eqn. 11 to choose a vector of weights for a hypothetical linear predictor and then simply set the integrator gain to the leading entry of this vector $g = w_1$. Simulations show that for common power-law noise processes, the resulting integrating servo performs almost as well as the optimal linear predictor, with a penalty of less than 10% in the prediction variance. The formalism we have developed can thus be used to optimise the parameters of a conventional integrating servo algorithm, without requiring any modifications to an already-running clock experiment. As we will see, this optimisation can be performed even without prior knowledge of the experiment's noise characteristics.

We quantify the scale of the servo's prediction errors by the dimensionless variance

$$v = \langle ((x - h) \omega T)^2 \rangle \quad (12)$$

of the phase accumulated in the Ramsey interrogation of duration T . The variance v plays an important role in determining both the robustness of the lock to atomic resonance and the achievable long-term stability (see Sec. IV), so that it is worth studying its behaviour. The solid lines in Fig. 2 illustrate the performance of linear predictors, based on simulations of clock operation with between 1 and 10^4 atoms in the reference. As a function of the choice of probe time T/Z , and thus of the ratio between LO noise and measurement noise, one can distinguish three qualitatively different regimes. In the limit of large atom numbers and long probe times, the simulations approach an N -independent limit

$$v = \xi \left(\frac{T}{Z} \right)^{2+\mu}. \quad (13)$$

This is the scaling one expects for the case where the LO noise completely dominates the measurement noise of the reference, given the postulated power-law scaling of the LO noise with exponent μ (c.f. Sec. I) and the definition of Z in Eqn. 3. The proportionality constant ξ can be estimated from the simulations or derived using the formulae in Appendix A, and varies between 1 (for a white-noise-dominated LO) and 2 (for a random-walk-noise-dominated LO). In the opposite limit of low atom number or short probe time, the (white) quantum projection noise dominates and the servo performance depends only on the number of measurements n which it averages in making its prediction, and thus

$$v \approx \frac{1}{Nn} \quad (14)$$

for sufficiently short probe times. For $n \gtrsim 20$ this limit has no impact on correctly optimised clock operation (see Sec. V). Between the two limits considered above there is a

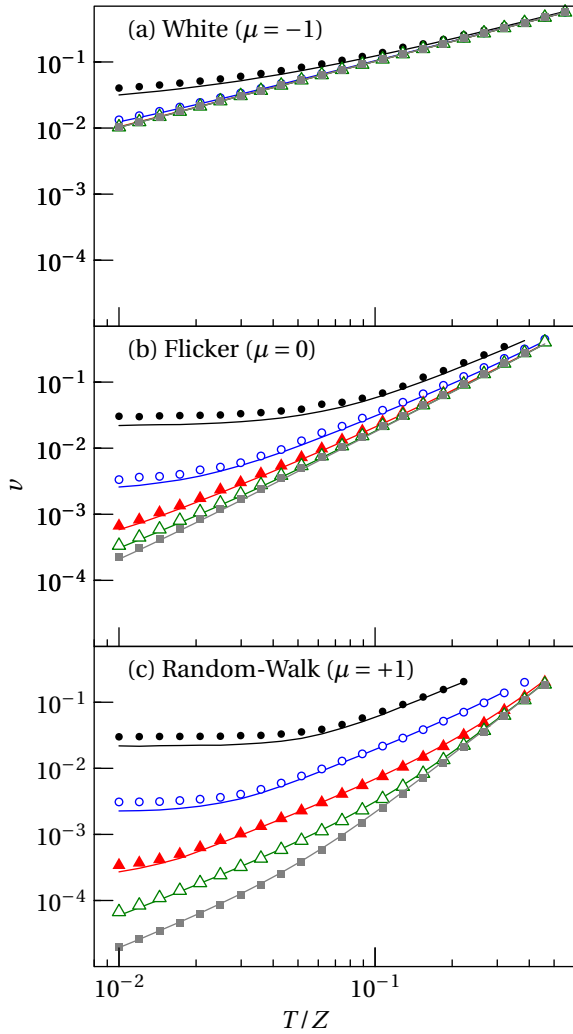


Figure 2. Variance of servo prediction errors, expressed as the variance ν of the phase accumulated in a Ramsey interrogation, as a function of probe time for a white-noise (top) flicker- (middle) or random-walk-limited (bottom) LO. The simulated clocks use Ramsey interrogation of 1 (solid black circles), 10 (open blue circles), 10^2 (solid red triangles), 10^3 (open green triangles) or 10^4 (solid grey squares) uncorrelated atoms. Symbols show the performance of integrating controllers optimised without knowledge of the LO noise as discussed in Sec. III. Solid lines show the performance of linear predictors using the last $n = 50$ frequency estimates, designed with knowledge of the LO noise properties as discussed in Sec. II.

trade-off between averaging many measurements to reduce the impact of measurement noise and considering only the recent measurements most relevant to the LO's current frequency. We know of no simple, accurate expression for the achievable servo performance in this intermediate regime, but the rough scaling that one would expect from the aforementioned trade-off

$$\nu \propto \frac{1}{\sqrt{N}} \left(\frac{T}{Z} \right)^{1+\mu/2} \quad (15)$$

does hold in simulations.

So far, we have discussed the simple integrator and its generalisations. Practical frequency standards, however, must use a double-integrator to correct for steady drifts in the LO frequency [15]. With the addition of the second integrator, the servo predictions become

$$h_0 = h_1 + g e_1 + g_2 \sum_k e_k \quad (16)$$

or

$$= \sum_k w_k y_k + g_2 \sum_k e_k. \quad (17)$$

Aside from its role in suppressing steady-state frequency errors with drifting LOs, the second integrator is necessary for the servo to have enough low-frequency gain to attain the projection noise limit in many-atom clocks, a point to which we will return in Sec. VII. However, as long as its gain g_2 is chosen low enough to avoid servo oscillations, the additional integrator has only a negligible impact on the variance of the prediction errors². The controller can thus be designed by optimising a simple integrator or linear predictor as discussed above, and then adding the drift-correction integrator with a gain $g_2 \ll w_1 = g$.

Besides minimising prediction variance, another desirable feature in practical servo controllers is *robustness*, the quality of remaining locked to the (correct) atomic resonance for long periods. In principle the two qualities are distinct, but we find empirically that for well-optimised servos they are tightly coupled. In simulations of clocks with a wide range of atom numbers (i.e. reference signal-to-noise ratios), we find that the rate at which a clock hops to different Ramsey fringes depends, for a given LO and a fully optimised servo, only on the prediction variance ν . Suboptimal servos (such as integrators with incorrectly chosen gain) have both greater prediction variance ν and a higher rate of fringe hops for a given ν , so that they are less robust as well as noisier. We conjecture that the best servos are simultaneously the most robust and the least noisy, so that there is no need to choose between the two qualities provided that one can, in fact, find this optimal servo design.

III. ON-LINE SERVO OPTIMISATION AND NOISE CHARACTERISATION

In practice, the noise spectrum of the LO may not be known accurately. A significant benefit of the formalism presented in the previous section is that it allows one to optimise the servo controller without prior knowledge of the

² This is best understood by considering the action of the servo in the frequency domain: the variance of the prediction errors depends on the feedback gain at a frequency corresponding to the clock cycle rate, whereas the second integrator only contributes feedback gain at much lower frequencies. The effect of the second servo is visible in the correlations between prediction errors, not in their variance.

LO noise properties. This is possible because the definition of \mathbf{C} in Eqn. 10 involves only the *estimated* LO frequency error in each clock cycle, which is routinely recorded in normal clock operation³. As a demonstration of such optimisation, we have run clock simulations with integrating servos whose gains were chosen, without knowledge of the true LO noise, by the following empirical procedure:

1. Start by setting the gain to $g = 0.2$, an arbitrary but reasonable initial value chosen to allow reliable, if suboptimal, clock operation under a wide range of conditions.
2. Simulate the clock for 10^4 cycles, corresponding to a few hours of operation for a typical contemporary frequency standard.
3. Compute \mathbf{C} according to Eqn. 10 and thence the vector \mathbf{w} of optimal weights. Set $g = w_1$ ⁴.
4. Simulate the clock with the newly optimised servo and reoptimise, repeating as necessary.

Even when the servo gain is initially chosen blindly, we find that five rounds of optimisation suffice for the gain g to converge to a value that yields performance indistinguishable from that of an integrator designed with full knowledge of the LO noise spectrum. Under more realistic conditions, where the initial choice of servo parameters reflects some prior knowledge of the LO performance, the optimisation could be performed much more quickly. The symbols in Fig. 2 show the prediction variance of such empirically-optimised integrators, which can be compared to the performance of optimal linear predictors shown as solid lines. For clocks operated near their optimal probe times (to be discussed in Sec. IV), the difference in prediction variance is less than 10%. Thus, it is possible to develop controllers that take full advantage of the time correlations in the LO noise even without independent knowledge of those correlations.

Unfortunately, it is not always possible to verify the servo performance directly, because the observed variance of the error signal e contains contributions both from the servo prediction error and from measurement noise of the reference. For a single-atom clock this problem is insurmountable: the observed fluctuations of a binary error signal must correspond to quantum projection noise, independent of servo performance. For a many-atom clock where the detection noise is well-characterised it is possible to measure the servo prediction variance as an increase in the fluctuations of the error signal, but the resulting estimates are

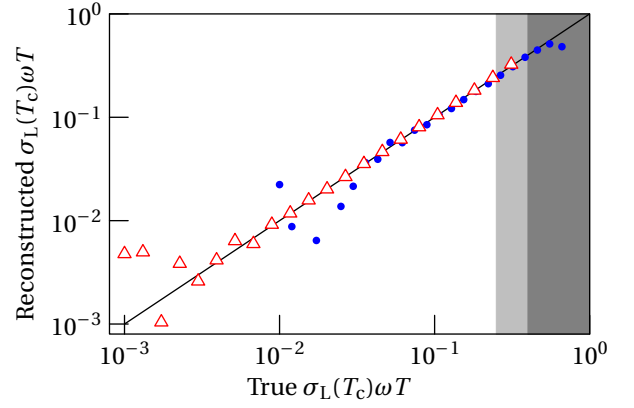


Figure 3. Normalised Allan deviation of flicker (blue circles) or random-walk (red triangles) LO noise reconstructed from correlations in the error signal of a single-atom clock as a function of the actual noise level. Black line marks correct reconstruction. Light and dark grey regions mark the noise levels at which the clock servo hops between Ramsey fringes for random walk and flicker noise respectively.

generally optimistic. As discussed in Sec. IV, even in a correctly optimised clock there will be unavoidable ambiguities in interpreting the error signal (e.g. 2π phase slips in Ramsey interrogation) and the resulting measurement errors contribute to the servo prediction variance without being observable in the experimentally recorded measurement data.

One can, however, use the correlation matrix \mathbf{C} estimated during clock operation to partially characterise the LO. Although white noise of the LO is indistinguishable from measurement noise in the atomic reference, flicker or random-walk noise can produce detectable temporal correlations even when their contributions to the total measurement variance are small. By fitting the estimated \mathbf{C} to a linear combination

$$\mathbf{C} = \sigma_w^2(T_c)\mathbf{C}_w + \sigma_f^2\mathbf{C}_f + \sigma_r^2(T_c)\mathbf{C}_r \quad (18)$$

of the correlation matrices expected for white, flicker and random-walk noise, one can obtain estimates of the Allan variances associated with each class of noise process, which are simply the coefficients in the linear combination. Explicit expressions for the correlation matrices associated with arbitrary noise spectra are provided in Appendix. A. Fig. 3, for instance, shows the flicker and random-walk Allan deviations reconstructed in this fashion from a correlation matrix \mathbf{C} estimated from 2×10^6 cycles of operation of a simulated single-atom clock, normalised to quantum projection noise, as a function of the true level of the respective noise processes. Random-walk noise, whose correlations differ more strongly from those of the white measurement noise, is easier to detect, but as seen in Fig. 3 both flicker and random walk noise can be reliably estimated from levels too low to affect the clock's instability (variance less than 1% of projection noise) up to levels that would be unacceptably high in normal operation, when the clock servo is jumping

³ In some implementations, the estimated frequency might not be recorded as such, but it can be obtained by adding the servo predictions to the frequency error signal reported by the atomic reference.

⁴ In cases where the noise is nearly white, this can lead to a theoretically optimal but practically useless controller with near-zero gain. To avoid this problem, we impose a minimum gain of 0.04 to ensure a finite servo attack time. This bound is low enough that it does not affect the achievable clock stability.

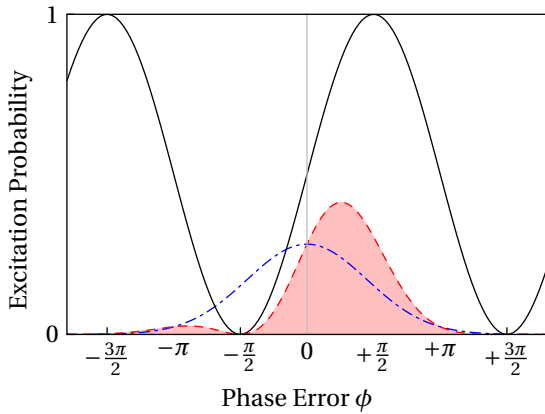


Figure 4. Schematic illustration of information gain in a single cycle of clock operation. Blue chain-dotted Gaussian: Prior knowledge of the LO frequency. Black solid sinusoid: Ramsey excitation spectrum. Red dashed line: Posterior knowledge of the LO frequency given that an atom was detected in the excited state. If no excitation is detected, then the posterior distribution is the mirror image about $\phi = 0$ of the one shown here.

between Ramsey fringes. Although this method provides much less detailed information on the LO noise spectrum than does the optical spectrum analyser of Ref. [34], it requires no measurements beyond those performed as part of the clock's normal operation. It can therefore be used to monitor the LO while the clock is running, even in a single-ion frequency standard, providing an early warning of performance degradations as well as information useful for the optimisation of interrogation parameters in the atomic reference (see Sec. V).

IV. IMPACT OF SERVO PERFORMANCE ON LONG-TERM STABILITY

Any measurement performed on a finite number of atoms can yield only a finite number of possible results. The optimisation of the measurement protocol thus involves a compromise between fine resolution over a narrow usable domain of LO frequencies and coarse frequency resolution over a broader domain. The best compromise depends on the range of frequencies which might plausibly have to be measured by the reference, i.e. on the variance of servo prediction errors. In this section we study this compromise, showing that it leads to a finite optimal probe time and an overall limit on the long-term stability of clocks with noisy LOs.

By way of illustration, consider Ramsey interrogation of a single atom, where the probability to find the atom in the excited state depends on the phase error $\phi = (x - h)\omega T$ as

$$P_e = \frac{1 + \sin \phi}{2}. \quad (19)$$

This excitation spectrum is shown as a solid line of Fig. 4. Let us assume that, before any measurement, our best knowledge of the (corrected) LO frequency is represented by the

chain-dotted distribution, corresponding to the distribution of LO prediction errors. Our best knowledge after a measurement on a single atom, in the event that it is found in the excited state, is shown by the dashed probability distribution. In this case we can be certain that the phase was not near $\phi = -\pi/2$ (since the excitation probability would then have been 0), and we can be reasonably confident that it lies in the region between 0 and π , but we cannot rule out that it lies near $-\pi$, where both the excitation probability and the prior probability distribution are non-negligible. Varying the probe time, and hence the spacing of the Ramsey fringes, involves a trade-off: increasing T narrows the main lobe of the posterior distribution, thanks to the steeper slope of the excitation signal, but also increases the weight of secondary lobes due to other fringes of the Ramsey spectrum.

To quantify this trade-off more formally, we consider Ramsey interrogation of N uncorrelated atoms, taking the prior distribution (chain-dotted curve in Fig. 4) to be a Gaussian of variance ν

$$P(\phi) = \frac{e^{-\frac{\phi^2}{2\nu}}}{\sqrt{2\pi\nu}}. \quad (20)$$

This distribution encodes, formally, all that is known about the LO frequency before the atomic measurement result becomes available. In a practical sense, it is the distribution of servo prediction errors: if the servo prediction were perfect ($h = x$) then the phase accumulated in the Ramsey interrogation would be zero. The assumption of normally distributed servo prediction errors is borne out by our simulations, which show that even with a single-atom reference and a binary error signal the skewness and excess kurtosis of the error distribution are no more than a few percent as long as the clock remains locked to the (correct) Ramsey fringe.

The reference does not, unfortunately, supply us with the expectation value P_e . Rather, the measurement yields a random fraction F of atoms detected in the excited state that fluctuates about the expectation value P_e due to measurement noise of the atomic reference. In the absence of technical noise on the reference signal, the variance of F is

$$\text{var}(F) = \left\langle (P_e - \langle P_e \rangle)^2 + \frac{P_e(1 - P_e)}{N} \right\rangle \quad (21)$$

$$= \frac{e^{-\nu} \sinh \nu}{4} + \frac{1 - e^{-\nu} \sinh \nu}{4N} \quad (22)$$

where the averages $\langle \cdot \rangle$ are taken over the prior distribution $P(\phi)$. The first term expresses the fluctuations in the measured excitation fraction due to actual changes in the ϕ -dependent excitation probability, while the second corresponds to quantum projection noise of the binomially-distributed excitation signal. The usefulness of the excitation fraction in estimating the frequency depends on the covariance of the two quantities

$$\text{cov}(\phi, F) = \langle \phi F \rangle - \langle \phi \rangle \langle F \rangle = \frac{\nu}{2} e^{-\frac{\nu}{2}} \quad (23)$$

which determines how much weight should be given to F in constructing the posterior estimate of the LO frequency.

Choosing the weight to minimise the variance v' of the error in this posterior estimate, we find (c.f. Appendix C):

$$v' = v - \frac{\text{cov}(\phi, F)^2}{\text{var}(F)} \quad (24)$$

$$= v \left(1 - \frac{Nv}{(N-1) \sinh v + e^v} \right). \quad (25)$$

This posterior variance combines information from the measurement with information that was known beforehand. In order to isolate the contribution of the former, we define an effective measurement variance v_m by

$$\frac{1}{v'} = \frac{1}{v} + \frac{1}{v_m}. \quad (26)$$

This is the usual relation for the variance of a (posterior) estimate obtained by an optimal linear combination of two *independent* pieces of information. v_m is thus the variance of a hypothetical measurement, one that could be interpreted without any prior knowledge, and which would reduce our uncertainty on the LO frequency as much as did the actual measurement. For the case we consider,

$$v_m = \frac{e^v}{N} + \left(1 - \frac{1}{N} \right) \sinh v - v \quad (27)$$

$$\approx \frac{1}{N} + \frac{v^2}{2N} + \frac{v^3}{6} + \dots \quad (28)$$

where, in the second line, we have expanded the effective measurement variance in powers of the prior variance. The first term is the conventional QPN on the measurement of the phase, valid when v is small and the corrected LO frequency is known *a priori* to be well centred on the Ramsey fringe. The third term reflects the additional uncertainty arising when the corrected LO frequency can lie outside the range where the reference produces a meaningful result. This term is independent of atom number, and dominates the effective measurement variance as v approaches 1.

Replacing the standard phase variance of Eqn. 1 by the effective measurement variance in Eqn. 2 yields a new prediction for clock stability in the limit of large averaging time τ , one that accounts for the effects of limited prior information in each interrogation:

$$\sigma_C(\tau) \rightarrow \frac{1}{\omega T} \sqrt{\frac{T_c}{\tau} \sqrt{\frac{e^v}{N} + \left(1 - \frac{1}{N} \right) \sinh v - v}}. \quad (29)$$

In Fig. 5 we compare this prediction with simulations for clocks with white-, flicker-, and random-walk-noise-limited LOs, using Ramsey interrogation of uncorrelated atoms with no dead time ($T_c = T$). Each point in these graphs is obtained from a simulation of 2×10^6 cycles of clock operation. The Allan deviation is computed for a time τ long enough that the instability has reached the asymptotic $1/\sqrt{\tau}$ regime (corresponding to 2×10^4 cycles of clock operation), then rescaled to a fixed noise level $\sigma_L(Z)$ to obtain a dimensionless result that is comparable across systems. The graphs thus show the

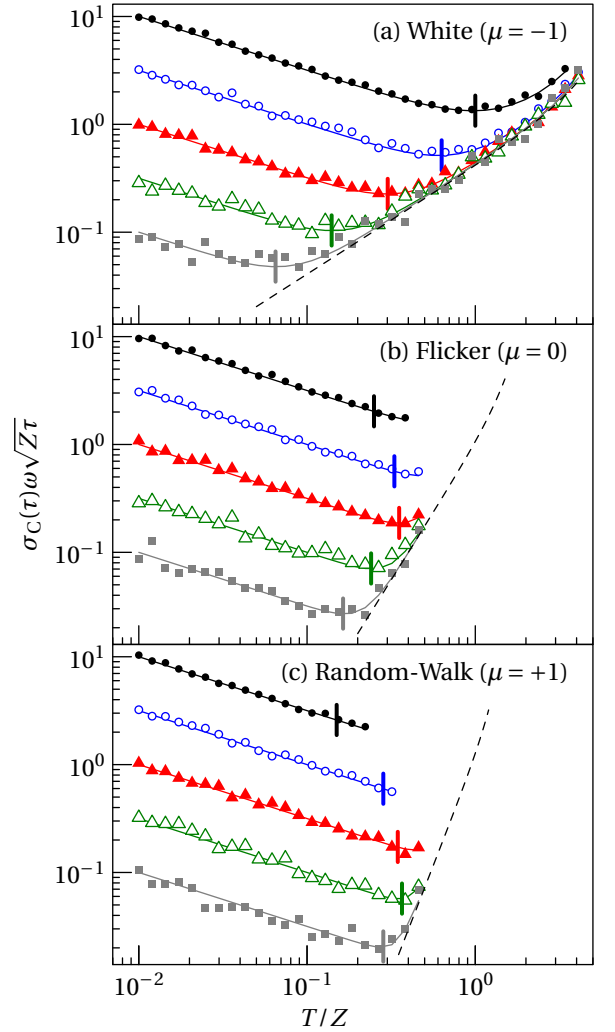


Figure 5. Long-term instability as a function of probe time for clocks using Ramsey interrogation of 1 (solid black circles), 10 (open blue circles), 10^2 (solid red triangles), 10^3 (open green triangles) or 10^4 (solid grey squares) uncorrelated atoms. Solid lines show the prediction of Eqn. 29, without free parameters. Vertical bars mark the recommended interrogation time given in Table II. Dashed line marks the limit for perfect phase estimation with no projection noise. The three graphs are, from top to bottom, for a white-, flicker-, or random-walk-dominated LO.

achievable long-term instability as a function of the choice of probe time.

When the probe time is short, the Ramsey fringe is broad and v is small, the instability improves as $1/\sqrt{T}$ as conventionally expected. The improvement with increasing probe time stops either when the additional v -dependent terms in Eqn. 29 grow important or when the servo can no longer reliably lock the LO to the reference transition: the curves in Fig. 5 end when the fringe-hop rate reaches 1 per 2 million cycles. As N increases, quantum projection noise is reduced relative to LO noise and it becomes advantageous to reduce the probe time so as to be less sensitive to the latter. Thus, the optimal probe time gets shorter with increasing

LO Noise Type	Asymptotic Scaling of $\sigma_C(\tau)\omega\sqrt{Z\tau}$
White	$\propto N^{-1/3}$
Flicker	$\propto N^{-5/12}$
Rnd. Walk	$\propto N^{-4/9}$

Table I. Asymptotic scaling of LO-limited clock instability with atom number. The scaling differs from the conventional $N^{-1/2}$ QPN limit because the optimum probe time decreases with increasing atom number.

atom number, and the fully optimised clock instability does not scale as $N^{-1/2}$. The asymptotic scaling with N is given in Table I. For white noise, the most extreme case, the optimal probe time scales as $N^{-1/3}$ in the large- N limit, leading to a $N^{-1/3}$ scaling of the long-term Allan deviation. For flicker or random-walk noise, ν falls off more steeply as the probe time is shortened (see Eqn. 13), so that the optimal probe time is less sensitive to atom number and a scaling closer to the conventional QPN limit is obtained. In the absence of projection noise, i.e. in the limit $N \rightarrow \infty$, the servo performance limit of Eqn. 13 combines with Eqn. 29 to yield a general measurement-noise-independent limit on clock instability, which we plot as dashed lines in Fig. 5. This limit arises solely from the unpredictability of the LO noise and from the finite domain over which the Ramsey error signal can be unambiguously interpreted.

Strictly speaking, Eqn. 24 holds only if the estimated frequency error is a linear function of the measured excitation fraction. Since the excitation probability is a non-linear (e.g. sinusoidal) function of the LO frequency error, one might hope to do better than the estimated performance of Eqn. 29 by using a non-linear function to convert the excitation fraction to a frequency error estimate. Simulations show, however, that correcting for the curvature of the Ramsey fringe by estimating the accumulated phase as $\arcsin(2F-1)$ rather than simply $2F-1$ has no significant effect on ν or on the achievable long-term clock stability. One can understand this finding by noting that, when ν is large enough that the curvature within a single Ramsey fringe is significant, the effect of unavoidable ambiguities such as the secondary lobe in Fig. 4 is much larger and dominates the posterior variance.

V. GUIDELINES FOR INTERROGATION PARAMETERS

To choose the operating parameters for a clock, one can in general use the formalism of Sec. II to predict the servo error variance ν as a function of those operating parameters and Eqn. 29 to predict the resulting long-term stability, which can then be optimised. Table II, for example, provides recommended Ramsey interrogation times for clocks dominated by different types of power-law LO noise, expressed as multiples of Z . In the many-atom limit, the servo prediction errors become independent of the quantum projection noise and we can solve Eqn. 29 to

LO Noise	Safe T/Z	Asymptotic Optimum T/Z
White	–	$\min(N^{-1/5}, 1.4 N^{-1/3})$
Flicker	$0.4 - 0.15 N^{-1/3}$	$0.76 N^{-1/6}$
Rnd. Walk	$0.4 - 0.25 N^{-1/3}$	$0.79 N^{-1/9}$

Table II. Recommendations for the choice of Ramsey interrogation time. The last column gives the optimal probe time in the limit of many atoms. There is nothing to be gained by probing longer than this time. It may be necessary to use shorter probe times to avoid fringe hops; a suggested safe upper bound on the probe time is given in the second column.

obtain the asymptotically optimal probe time (last column of Table II). Increasing the probe time beyond this optimum always leads to an increase in effective measurement variance and long-term instability, and is of no practical interest. At small atom numbers, shorter probe times are required to keep the servo controller robust against fringe hops. The purely phenomenological bound in the second column of Table II is chosen to be slightly shorter than the time for which we observe fringe-hops at a rate of 1 per million simulated clock cycles, with a 20% safety margin. Our choice of maximum acceptable fringe-hop rate, corresponding to a requirement that the clock remain locked to the correct fringe for a few days, is arbitrary, but as the onset of fringe-hopping is extremely steep (the fringe-hop rate in simulations increases by two to three orders of magnitude when the probe time is doubled), the maximum safe probe time is only weakly dependent on this choice of threshold. A full optimisation of all common probe protocols in the presence of realistic experimental imperfections is beyond the scope of this work, but we expect qualitatively similar behaviour from Rabi or hyper-Ramsey probing, with somewhat longer optimal probe times and slightly degraded instability due to the increased width of the observed atomic resonance in these schemes. Conversely, we expect that clocks with significant dead time in their operating cycle will need to use somewhat shorter probe times to compensate for the servo's inability to correct unobserved LO frequency fluctuations, which will lead to a ν higher than in our dead-time-free simulations.

VI. CONSTRAINTS ON THE BENEFITS OF ENTANGLED ATOMIC REFERENCES

The arguments developed in the preceding sections also apply to certain Ramsey-like protocols using entangled states in the atomic reference, provided that LO noise is the limiting form of decoherence. For instance, the scheme proposed in Ref. [35] and demonstrated in Refs. [36–38], employing N atoms in a maximally correlated state $|\psi\rangle = (|g\rangle^{\otimes N} + |e\rangle^{\otimes N})/\sqrt{2}$, where $|g\rangle$ and $|e\rangle$ are the atomic eigenstates, is fully equivalent to Ramsey interrogation of a single atom with an N -fold enhanced transition frequency by the corresponding harmonic of the LO radiation. Now

the long-term instability of a single-atom frequency standard can be expressed as

$$\sigma_C(\tau) = \frac{s}{\omega\sqrt{Z\tau}}, \quad (30)$$

with s a dimensionless constant of order unity encoding the choice of probe time T/Z and the additional contribution of LO noise at this probe time. The long-term instability of the clock using maximally-correlated atoms thus becomes

$$\sigma_C(\tau) = \frac{s}{N\omega\sqrt{Z_N\tau}}, \quad (31)$$

where Z_N is the noise timescale for the N -fold frequency-multiplied LO:

$$\sigma_L(Z_N)N\omega Z_N = 1 \text{ rad}. \quad (32)$$

In the dead-time-free limit $Z_{Nc} = Z_N$, one can compare Eqn. 32 with Eqn. 3 and find

$$\frac{Z_N}{Z} = N^{\frac{-2}{2+\mu}}, \quad (33)$$

with μ again describing the time-dependence of the LO Allan variance (see Sec. I). The entangled clock's long-term instability thus scales as

$$\sigma_C(\tau) = \frac{s}{\omega\sqrt{Z\tau}} \frac{1}{N^{\frac{1+\mu}{2+\mu}}}. \quad (34)$$

Thus, if the clock stability is limited by white LO noise ($\mu = -1$), a reference using a maximally-correlated state of N atoms performs no better than a reference using a *single* atom, and is in fact *worse* than a reference using uncorrelated interrogation of the N atoms. This scaling has been observed experimentally for correlated magnetic field noise in a 14 ion GHZ state [38]. For flicker-floor LO noise, the Allan deviation improves as $N^{-1/2}$ with N maximally-entangled atoms, very slightly better than the asymptotic $N^{-5/12}$ scaling achievable without entanglement, but worse than the scaling achieved with unentangled atoms for $N < 10^2$. It is only for random-walk LO noise that maximally entangled states offer measurable benefits, with an $N^{-2/3}$ scaling of the long-term Allan deviation. We illustrate these scalings in Fig. 6, which plots the Allan deviation spectrum recorded in simulations of fully optimised clocks using either 100 uncorrelated atoms or a 100-atom maximally-correlated state for all three LO noise types.

Maximally entangled states reduce the signal-to-noise ratio of a measurement on N atoms to that of a single qubit, providing less new information per measurement but accelerating the clock cycle so that more measurements can be averaged. That is why their use is advantageous with random-walk LO noise, when fast measurements can take advantage of the reduced LO noise at short time scales. Other approaches to the use of entanglement in atomic references, such as spin squeezing [24–26, 39–41], focus instead on improving the signal-to-noise ratio of the measurement, and thus increasing the amount of new information obtained in each interrogation. The error signal produced in such schemes has the same periodic ambiguities

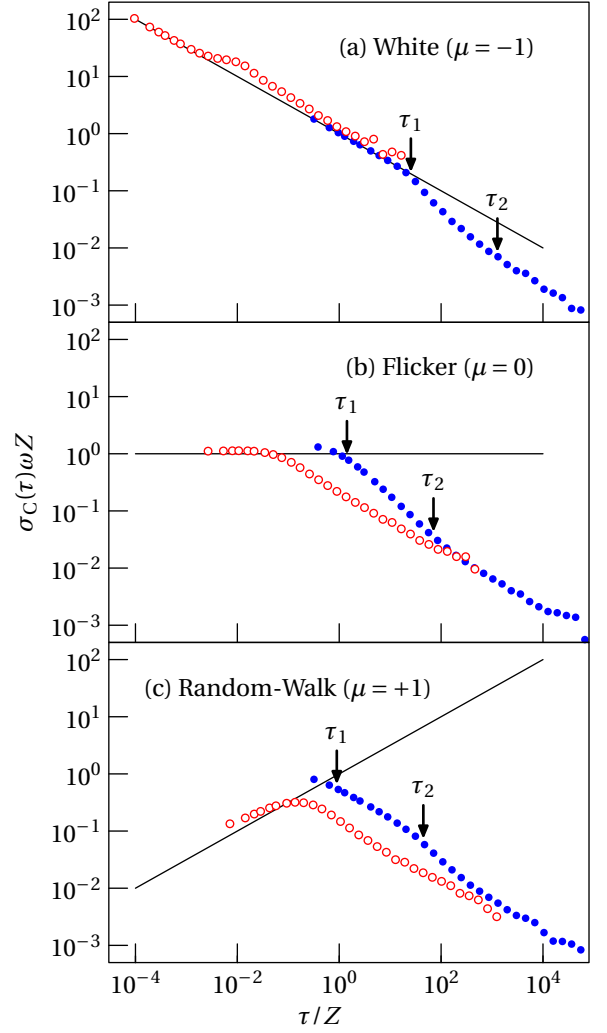


Figure 6. Simulated Allan deviation as a function of averaging time for optimised clocks using either 100 uncorrelated atoms (solid blue circles) or a maximally-correlated state of 100 atoms (open red circles). The latter must operate with shorter interrogation times, and thus the corresponding curves start earlier. The black line marks the Allan deviation of the free-running LO. The three graphs are, from top to bottom, for a white-, flicker-, or random-walk-dominated LO. Arrows mark the two integrator time constants for the case of uncorrelated atoms, as discussed in Sec. VII.

as in Ramsey interrogation of uncorrelated atoms, and so they are subject to the same projection-noise-independent $\sinh v - v \approx v^3/6$ limit on their effective measurement variance (c.f. Eqn. 28). The additional noise introduced when servo prediction errors allow the anti-squeezed quadrature to contaminate the measurement result [42] can in principle be eliminated by a suitable readout procedure [27] in which case we expect such interrogation protocols to offer benefits comparable to suppressing the projection noise by increasing atom number, even with white- or flicker-floor limited LOs.

VII. SOME REMARKS ON SHORT-TERM INSTABILITY

So far we have focused on the long-term instability of the clock once it reaches the asymptotic $\sigma_C \propto 1/\sqrt{\tau}$ regime, without considering the averaging time required to reach this regime. In general, a clock reaches its asymptotic instability when the fluctuations in the frequency of the output signal are dominated by the measurement noise of the atomic reference. For single-ion clocks in which the signal-to-noise ratio of the atomic measurements is no better than 1, this condition is reached at the servo attack time τ_1 , as soon as the output signal of the clock stops following the free-running LO and is locked to the noisy signal from the atomic reference. However, clocks using many atoms have a much higher signal-to-noise ratio, i.e. the resolution of the error signal from their atomic reference is much finer than the LO frequency fluctuations that they can reliably measure. In such clocks the optimal probe times are long enough that the quantum projection noise is well below the LO-limited short-term instability. In order to reach the asymptotic regime they must initially average down faster than $1/\sqrt{\tau}$. This is possible provided that the servo has enough gain to suppress the measured LO frequency fluctuations. A single integrator can suppress measured LO fluctuations by a factor $\sim \tau/\tau_1$ in the standard deviation⁵, so that the clock instability initially averages down as $\tau^{-3/2}$ or τ^{-1} for white or flicker LO noise respectively. This is fast enough to reach the measurement-noise limited regime in an averaging time of roughly $\sqrt{N}\tau_1$ or $N\tau_1$ for a white-noise-limited or flicker-floor limited LO respectively, as seen in the first two graphs of Fig. 6. However, a single integrator can only suppress random-walk LO noise to a level that scales as $\tau^{-1/2}$, which will not catch up with the measurement noise limit which is averaging down at the same rate. A many-atom clock using only a single integrator would thus be forever limited by the finite gain of the servo rather than by the noise of the atomic measurements. It is only when a second integrator allows the servo to suppress noise by an additional factor of τ/τ_2 that the clock instability can average down as $\tau^{-3/2}$ until it reaches the measurement noise limit in a time $\sqrt{N}\tau_2$. The third graph of Fig. 6 illustrates this behaviour, with the uncorrelated 100-atom clock initially averaging down at the servo-limited rate of $\tau^{-1/2}$ until around $\tau_2 \approx 45Z$. The second integrator then allows the instability to catch up with the lower-lying asymptotic noise limit, which it reaches around $\tau \approx 400Z$. It is interesting to note that clocks using maximally-correlated states, because they behave like single-atom clocks and are always measurement-noise limited, would have an advantage in short-term instability even when their long-term instability is little better than that of a clock with uncorrelated atoms (lower two graphs of Fig. 6). This observation mirrors, in a simpler setting, the finding of Ref. [20].

⁵ This is the time-domain equivalent of the observation that an integrating servo suppresses noise power at a frequency f by a factor proportional to f^2 .

Thus the second integrator in a clock servo, beyond its role in correcting for linear drifts, is also needed to suppress random-walk noise of the LO in many-atom clocks. It is desirable to set the gain g_2 of this drift-correction integrator as high as possible, in order to reach the asymptotic instability in a reasonable time. However, it must not be so high that it induces oscillations in the lock. With a conventional two-stage integrating servo, the ratio of the two gains must be no more than a few percent (we use $g_2 = g/50$ in our simulations). Linear predictors optimised as in Sec. II are somewhat more robust against oscillations, and can be operated with higher gain $g_2 = w_1/10$ for the drift-correction integrator.

Note that when post-processing measurement results to generate a virtual “paper” clock signal, the causality requirements which limit the gain of the servo during physical clock operation no longer apply. Thus, while the long-term stability limits discussed in Sec. IV hold equally for physical and paper clocks because they arise from limits on the noise of the atomic reference, the short-term stability limits discussed in this section can be avoided entirely in paper clocks and frequency ratio measurements, where the LO frequency fluctuations can always be corrected as well as they can be measured.

VIII. OUTLOOK

In this work we have studied the effects of LO noise on frequency standards that monitor the LO frequency using a single ensemble of atoms periodically interrogated according to a fixed protocol and that correct the measured frequency fluctuations using a linear prediction formula. Most current optical atomic clocks fit this description and can, without hardware modifications, use the framework presented here to identify and approach the stability limit imposed by their LO performance. The interrogation times we recommend are specific to dead-time-free Ramsey interrogation, but qualitatively similar results for other (Rabi, hyper-Ramsey) protocols can be found by the same arguments, since our treatment of the servo is protocol-independent and since all interrogation protocols face the same trade-off between measurement resolution and unambiguous measurement domain. Within this framework, the most promising approaches to improving long-term clock instability (besides improving LO performance) seem to be those that improve the dynamic range of atomic measurements (such as spin squeezing), whereas methods which attempt to make faster measurements with poor dynamic range (such as spectroscopy with maximally-correlated states) have been shown to offer modest or no benefits for realistic LO noise spectra.

There are, however, many architectures for frequency standards that do not fit the framework presented here, and it would be interesting to consider which of them can overcome the limits we have identified. The simplest extension to implement would be the use of non-linear prediction algorithms, which might improve the robustness of the servo,

allowing longer probe times and better stability at small atom number. We expect that the performance of such algorithms would still be subject to the measurement-noise-independent limit of Eqn. 13, so that they are unlikely to offer more than a modest constant-factor stability improvement in the large- N limit.

Proposed multi-ensemble or cascaded clocks [19, 20] circumvent the limits we have discussed here by monitoring the LO noise with several different atomic references with progressively finer resolution. References with a broad domain of useful frequencies provide coarse-resolution results sufficient to narrow the prior ν for other, finer-resolution references. The analysis we have presented here applies directly to the first (coarsest) reference in the cascade, and the resulting stabilised signal can then be treated as an effective LO used by the next reference in the ensemble, thus proceeding step-by-step down the cascade. However, even if our analysis is locally valid for every reference treated individually, the overall behaviour of such a multi-ensemble system may be qualitatively different than would be naively expected from the single-ensemble analysis [19, 20, 22].

Finally, it would be interesting to make an analogous study for continuously-interrogated atomic references [43–50]. Such systems, whether based on continuous spectroscopic observation of an atomic sample or on direct lasing on the clock transition (“active optical clock”) [46, 47] face a conceptually similar trade-off between suppressing the noise of the atomic signal (driving the system weakly to minimise the disturbance to the atoms) and suppressing classical fluctuations in the probe laser, cavity mirrors, etc. (driving the system strongly to gain information quickly and maximise the useful feedback bandwidth). Thus, the stability of these superficially different systems may depend on the noise of their classical components and on the size of the atomic sample in ways qualitatively similar to those we have examined here.

We thank S. King and R. Demkowicz-Dobrzański for stimulating discussions. I.D.L. acknowledges a fellowship from the Alexander von Humboldt foundation. The work presented here was partly supported through the project EMPIR 15SIB03 OC18. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme. We acknowledge support from the DFG through CRC 1128 (geo-Q), project A03 and CRC 1227 (DQ-mat), project B03.

Appendix A: Fluctuation correlation matrices C for common noise processes

In order to reconstruct LO noise properties from the experimentally observed correlation matrix C , it is helpful to have explicit expressions for the correlations induced by common power-law noise processes, which we summarise here.

Consider a continuous noisy process $y(t)$ with one-sided

power spectral density $S(f)$, whose autocorrelation reads

$$\langle y(t)y(t+\tau) \rangle = \int_0^\infty S(f) \cos(2\pi f\tau) df. \quad (A1)$$

If we associate the discrete estimates y_j with time averages over a clock cycle of duration T_c such as would be measured by a perfect classical frequency counter,

$$y_j = \frac{1}{T_c} \int_{-jT_c}^{(1-j)T_c} y(t) dt, \quad (A2)$$

then the definition of Eqn. 10 reduces to

$$C_{jk} = \frac{4}{\pi T_c} \int_0^\infty \frac{S(u/\pi T_c)}{u^2} \sin^2(u) \sin(uj) \sin(uk) \times \cos(u(j-k)) du \quad (A3)$$

where u is a dimensionless dummy integration variable.

For a white-noise process ($S(f) \propto f^0$) of Allan deviation σ_w , we find

$$C = \sigma_w^2(T_c) C_w \quad (A4)$$

$$(C_w)_{jk} = 1 + \delta_{jk} \quad (A5)$$

$$C_w = \begin{pmatrix} 2 & 1 & 1 & 1 & \cdots \\ 1 & 2 & 1 & 1 & \cdots \\ 1 & 1 & 2 & 1 & \cdots \\ 1 & 1 & 1 & 2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (A6)$$

as expected. Detection noise of the atomic reference, though not associated with the continuous frequency fluctuations of the LO, yields the same white-noise correlation matrix.

For a flicker-noise process ($S(f) \propto f^{-1}$) of Allan deviation σ_f ,

$$C = \sigma_f^2 C_f \quad (A7)$$

$$(C_f)_{jk} = \mathcal{D}(|j-k|) - \mathcal{D}(j) - \mathcal{D}(k) \quad (A8)$$

$$C_f \approx \begin{pmatrix} 2 & 1.57 & 1.30 & 1.21 & \cdots \\ 1.57 & 3.13 & 2.43 & 2.08 & \cdots \\ 1.30 & 2.43 & 3.74 & 2.95 & \cdots \\ 1.21 & 2.08 & 2.95 & 4.16 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (A9)$$

where the auxiliary function \mathcal{D} is defined as

$$\mathcal{D}(n) = \frac{2\mathcal{L}(n) - \mathcal{L}(n-1) - \mathcal{L}(n+1)}{4} \quad (A10)$$

with

$$\mathcal{L}(n) = \begin{cases} n^2 \log_2 n & n > 1 \\ 0 & \text{otherwise} \end{cases} \quad (A11)$$

For random-walk process ($S(f) \propto f^{-2}$) of Allan deviation σ_r ,

$$C = \sigma_r^2(T_c) C_r \quad (\text{A12})$$

$$(C_r)_{jk} = 3 \min(j, k) - \frac{1 + \delta_{jk}}{2} \quad (\text{A13})$$

$$C_r = \begin{pmatrix} 2 & \frac{5}{2} & \frac{5}{2} & \frac{5}{2} & \dots \\ \frac{5}{2} & 5 & \frac{11}{2} & \frac{11}{2} & \dots \\ \frac{5}{2} & \frac{11}{2} & 8 & \frac{17}{2} & \dots \\ \frac{5}{2} & \frac{11}{2} & \frac{17}{2} & 11 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (\text{A14})$$

Note that C_{11} is, by definition, twice the single-cycle Allan variance for the noise process under consideration. Also note that, while Eqn. A3 is valid only for perfect frequency-counting or dead-time free Ramsey interrogation, Eqns. A4 through A14, expressed in terms of observed Allan variances, are valid for arbitrary noisy time series of the specified power-law noise type, including frequency estimates made with arbitrary measurement protocols that may include dead time.

Appendix B: Optimal integrator gain with known noise

In simulations or in experiments where the LO is well-characterised, it can be helpful to have an explicit formula for the optimal integrator gain in terms of known LO noise parameters. To derive such a formula, we start by noting that for a pure white-noise process, the variance of prediction errors for a simple integrator with gain g is

$$\sigma_w^2(T_c) \frac{2}{2-g}, \quad (\text{B1})$$

as can be shown by direct application of Eqn. 9 to the correlation matrix of Eqn. A6. Similarly, for a pure random-walk process (and a hypothetical perfect classical reference), the variance of prediction errors would be

$$\sigma_r^2(T_c) \frac{3-g}{g(2-g)}. \quad (\text{B2})$$

We have not found a similarly simple formula for the case of flicker noise, but numerical studies shows that the phenomenological equation

$$\sigma_f^2(T_c) \frac{1.6 + 0.4g - \ln 4 \ln g}{2-g} \quad (\text{B3})$$

is accurate to within 2% for gains in the range of 10^{-3} to 1.

As the integrator is a linear controller, the mean-squared prediction error for a general noise process combining white, flicker and random-walk contributions is simply the sum of the three preceding expressions. Differentiating this sum with respect to g and imposing the condition $g > 0$ to

exclude servos that do nothing at all, we find that the optimum gain (the one which minimises the variance of prediction errors) must satisfy

$$-2g^2 \beta \ln 2 \ln g + (2 - \rho + (2.4 + \ln 4)\beta)g^2 + (6\rho - \beta \ln 16)g - 6\rho = 0 \quad (\text{B4})$$

where $\beta = \sigma_f^2(T_c)/\sigma_w^2(T_c)$ and $\rho = \sigma_r^2(T_c)/\sigma_w^2(T_c)$ are the flicker-floor and random-walk Allan variances at one cycle, normalised to the total variance of all white noise contributions (including reference noise). The first term on the left-hand side is always small and can be safely neglected, leaving us with a quadratic equation in g . Solving this equation yields a prescription for the gain

$$g \approx \frac{\beta \ln 4 - 3\rho + \sqrt{(\beta \ln 4 - 3\rho)^2 + 6\rho a}}{a} \quad (\text{B5})$$

where we have introduced the usual auxiliary quantity $a = 2 + (2.4 + \ln 4)\beta - \rho$. Although approximate, Eqn. B5 yields integrating controllers whose prediction error variance comes within 1% of that of the best numerically optimised integrators.

Note that for systems where the LO random-walk noise dominates all other noise sources (large ρ), Eqn. B5 can recommend an integrator gain in excess of 1. This may seem surprising, but is correct. If, for instance, the last measured frequency was higher than the corresponding prediction, it is likely that the LO frequency was random-walking upwards during the last measurement, and thus that the frequency at the end of the measurement was higher than the average frequency during the measurement. The apparent over-correction implied by $g > 1$ is accounting for this difference. The integrating servo is still stable as long as $g < 2$.

Also note that, in the absence of flicker- or random-walk noise, the theoretically optimal integrator has a vanishingly small gain in order to average the white noise down as far as possible. In practice, one should enforce a minimum acceptable gain, in order to keep the servo time constant from growing unreasonable. All our simulations enforce $g > 0.04$, i.e. a servo attack time no longer than 25 clock cycles, though this bound is usually only relevant for unrealistic LO noise models or when the interrogation time of the atomic reference is too short.

Appendix C: Posterior variance

Equation 24 gives the variance v' of errors in the posterior estimate of the LO frequency as a function of the variance v of the servo prediction errors in the prediction and of statistical properties of the measurement signal F . Here we review the derivation of this posterior variance.

We wish to construct an estimate ψ of the unknown phase ϕ corresponding to the error of the servo's prediction. We take our estimate to be a linear combination of two (not necessarily independent) pieces of information: the expectation value $\langle \phi \rangle$, which captures information available before

the measurement result is revealed, and the measurement result itself. For later convenience, we express the measurement result as a deviation from the expectation value $\langle F \rangle$ of the measurement signal F . Thus:

$$\psi = \alpha \langle \phi \rangle + \beta (F - \langle F \rangle). \quad (C1)$$

As in Sec. IV, we define ϕ such that $\langle \phi \rangle = 0$. The mean squared error ν' which we wish to minimise is then

$$\nu' = \langle (\psi - \phi)^2 \rangle = \langle [\beta (F - \langle F \rangle) - \phi]^2 \rangle \quad (C2)$$

$$= \beta^2 \langle (F - \langle F \rangle)^2 \rangle - 2\beta \langle (F - \langle F \rangle)\phi \rangle + \langle \phi^2 \rangle \quad (C3)$$

$$= \beta^2 \text{var}(F) - 2\beta \text{cov}(\phi, F) + \nu \quad (C4)$$

The minimum is attained for

$$\beta = \frac{\text{cov}(\phi, F)}{\text{var}(F)}. \quad (C5)$$

Note that the weight β given to the latest measurement result in estimating the frequency of the LO decreases as the measurement becomes noisier (i.e. as $\text{var}(F)$ grows) or as the measurement becomes less strongly correlated with the underlying phase ϕ that we wish to estimate (i.e. as $\text{cov}(\phi, F)$ shrinks). The minimum posterior variance given in Eqn. 24 is obtained directly upon substitution of the optimised weight β into Eqn. C4.

For simplicity, we restrict ourselves to a *linear* combination of the measurement signal F with the prior estimate $\langle \phi \rangle$, as this allows the results to be expressed entirely in terms of experimentally accessible (co)variances of noise distributions. The variance of the errors in non-linear estimators depends on higher-order moments of the noise distributions which are difficult to characterise experimentally. As argued at the end of Sec. IV, non-linear estimators are empirically unnecessary, at least for simple Ramsey-like protocols.

-
- [1] A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, and P. O. Schmidt, *Rev. Mod. Phys.* **87**, 637 (2015).
 - [2] N. Poli, C. W. Oates, P. Gill, and G. M. Tino, *Riv. Nuovo Cimento* **36**, 555 (2013).
 - [3] M. Vermeer, Reports of the Finnish Geodetic Institute **83(2)**, 1 (1983).
 - [4] A. Bjerhammar, *Bull. Geodesique* **59**, 207 (1985).
 - [5] P. Delva and J. Lodewyck, “Atomic clocks: New prospects in metrology and geodesy,” arXiv:1308.6766 (2013).
 - [6] C. Lisdat, G. Grosche, N. Quintin, C. Shi, S. M. F. Raupach, C. Grebing, D. Nicolodi, E. Stefani, A. Al-Masoudi, S. Dörscher, S. Häfner, J.-L. Robyr, N. Chiodo, S. Bilicki, E. Bookjans, A. Koczwar, S. Koke, A. Kuhl, F. Wiotte, F. Meynadier, E. Camisard, M. Abgrall, M. Lours, T. Legero, H. Schnatz, U. Sterr, H. Denker, C. Chardonnet, Y. Le Coq, G. Santarelli, A. Amy-Klein, R. Le Targat, J. Lodewyck, O. Lopez, and P.-E. Pottie, *Nat Commun* **7**, 12443 (2016).
 - [7] A. Derevianko and M. Pospelov, *Nat. Phys.* **10**, 933 (2014).
 - [8] K. Van Tilburg, N. Leefer, L. Bougas, and D. Budker, *Phys. Rev. Lett.* **115**, 011802 (2015).
 - [9] Y. V. Stadnik and V. V. Flambaum, *Phys. Rev. A* **94**, 022111 (2016).
 - [10] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, and D. J. Wineland, *Phys. Rev. A* **47**, 3554 (1993).
 - [11] F. Riehle, *Frequency Standards: Basics and Applications* (Wiley-VCH, 2004).
 - [12] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *Phys. Rev. Lett.* **79**, 3865 (1997).
 - [13] E. Riis and A. G. Sinclair, *J. Phys. B: At. Mol. Opt. Phys.* **37**, 4719 (2004).
 - [14] C. Champenois, M. Houssin, C. Lisowski, M. Knoop, G. Hagel, M. Vedel, and F. Vedel, *Phys. Lett. A* **331**, 298 (2004).
 - [15] E. Peik, T. Schneider, and C. Tamm, *J. Phys. B* **39**, 145 (2006).
 - [16] G. J. Dick, in *Proc. 19th Precise Time and Time Interval* (Rendondo Beach, Ca, 1987) pp. 133–147.
 - [17] V. Bužek, R. Derka, and S. Massar, *Phys. Rev. Lett.* **82**, 2207 (1999).
 - [18] G. W. Biedermann, K. Takase, X. Wu, L. Deslauriers, S. Roy, and M. A. Kasevich, *Phys. Rev. Lett.* **111**, 170802 (2013).
 - [19] T. Rosenband and D. Leibbrandt, “Exponential scaling of clock stability with atom number,” arXiv:1303.6357v2 (2013).
 - [20] E. M. Kessler, P. Kómár, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, *Phys. Rev. Lett.* **112**, 190403 (2014).
 - [21] M. Mullan and E. Knill, *Phys. Rev. A* **90**, 042310 (2014).
 - [22] K. Chabuda, I. D. Leroux, and R. Demkowicz-Dobrzański, *New J. Phys.* **18**, 083035 (2016).
 - [23] M. Fraas, *Commun. Math. Phys.* **348**, 363 (2016).
 - [24] M. Kitagawa and M. Ueda, *Phys. Rev. A* **47**, 5138 (1993).
 - [25] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, *Phys. Rev. A* **46**, R6797 (1992).
 - [26] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen, *Phys. Rev. A* **50**, 67 (1994).
 - [27] J. Borregaard and A. S. Sørensen, *Phys. Rev. Lett.* **111**, 090801 (2013).
 - [28] N. Huntemann, B. Lipphardt, M. Okhapkin, C. Tamm, E. Peik, A. V. Taichenachev, and V. I. Yudin, *Phys. Rev. Lett.* **109**, 213002 (2012).
 - [29] V. I. Yudin, A. V. Taichenachev, C. W. Oates, Z. W. Barber, N. D. Lemke, A. D. Ludlow, U. Sterr, C. Lisdat, and F. Riehle, *Phys. Rev. A* **82**, 011804 (2010).
 - [30] N. Wiener, *The interpolation, extrapolation and smoothing of stationary time series*, Tech. Rep. Report of the Services 19, Research Project DIC-6037 (MIT, 1942).
 - [31] A. N. Kolmogorov, *Izv. Akad. Nauk SSSR Ser. Mat.* **5**, 3 (1941).
 - [32] P. P. Vaidyanathan, *The Theory of Linear Prediction* (Morgan & Claypool, 2008).
 - [33] E. H. Isakks, *An Introduction to Applied Geostatistics* (Oxford University Press, New York, 1989).
 - [34] M. Bishof, X. Zhang, M. J. Martin, and J. Ye, *Phys. Rev. Lett.* **111**, 093604 (2013).
 - [35] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, *Phys. Rev. A* **54**, R4649 (1996).
 - [36] V. Meyer, M. A. Rowe, D. Kielpinski, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, *Phys. Rev. Lett.* **86**, 5870 (2001).
 - [37] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland, *Nature* **438**, 639 (2005).

- [38] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, *Phys. Rev. Lett.* **106**, 130506 (2011).
- [39] I. D. Leroux, M. H. Schleier-Smith, and V. Vuletić, *Phys. Rev. Lett.* **104**, 250801 (2010).
- [40] A. Louchet-Chauvet, J. Appel, J. J. Renema, D. Oblak, N. Kjærgaard, and E. S. Polzik, *New J. Phys.* **12**, 065032 (2010).
- [41] O. Hosten, N. J. Engelsen, R. Krishnakumar, and M. A. Kasevich, *Nature (London)* **529**, 505 (2016).
- [42] A. André, A. S. Sørensen, and M. D. Lukin, *Phys. Rev. Lett.* **92**, 230801 (2004).
- [43] M. J. Martin, D. Meiser, J. W. Thomsen, J. Ye, and M. J. Holland, *Phys. Rev. A* **84**, 063813 (2011).
- [44] P. G. Westergaard, B. T. R. Christensen, D. Tieri, R. Martin, J. Cooper, M. Holland, J. Ye, and J. W. Thomsen, *Phys. Rev. Lett.* **114**, 093002 (2015).
- [45] B. T. R. Christensen, M. R. Henriksen, S. A. Schäffer, P. G. Westergaard, D. Tieri, J. Ye, M. J. Holland, and J. W. Thomsen, *Phys. Rev. A* **92**, 053820 (2015).
- [46] J. Chen, *Chinese Science Bulletin* **54**, 348 (2009).
- [47] D. Meiser, J. Ye, D. R. Carlson, and M. J. Holland, *Phys. Rev. Lett.* **102**, 163601 (2009).
- [48] J. G. Bohnet, Z. Chen, J. M. Weiner, D. Meiser, M. J. Holland, and J. K. Thompson, *Nature (London)* **484**, 78 (2012).
- [49] N. Shiga and M. Takeuchi, *New J. Phys.* **14**, 023034 (2012).
- [50] M. A. Norcia and J. K. Thompson, *Phys. Rev. X* **6**, 011025 (2016).